A commercial flight track database for upper tropospheric aircraft emission studies over the USA and southern Canada

DONALD P. GARBER, PATRICK MINNIS* and P. KAY COSTULIS

Science Directorate, NASA Langley Research Center, Hampton, Virginia, USA

Abstract

A new air traffic database over the contiguous United States of America (USA) and southern Canada has been developed from a commercially available real-time product for 2001–2003 for all non-military IFR flights above 25,000 ft. Both individual flight tracks and gridded spatially integrated flight legs are available. On average, approximately 27,350 high-altitude flights were recorded each day. The diurnal cycle of air traffic over the USA is characterized by a broad daytime maximum with a 0130-LT minimum and a mean day-night air traffic ratio of 2.4. Each week, the air traffic typically peaks on Thursday and Friday and drops to a low during Sunday with a range of 18 %. Flight density is greatest during late summer and least during winter. The database records the disruption of air traffic after the air traffic shutdown during September 2001. The dataset should be valuable for realistically simulating the atmospheric effects of aircraft in the upper troposphere.

Zusammenfassung

Ein neues Flugverkehrs-Kataster für den zusammenhängenden Teil der Vereinigten Staaten von Amerika (USA) wurde auf der Basis eines für die Jahre 2001 bis 2003 kommerziell verfügbaren Echtzeitprodukts für alle zivilen Flugbewegungen nach Instrumentenflugregeln (IFR) oberhalb von 25000 Fuß entwickelt. Das Kataster enthält sowohl Informationen über individuelle Flugrouten als auch räumlich über Gitterboxen integrierte Daten. Im Mittel wurden täglich rund 27350 hochfliegende Flugverbindungen registriert. Der Tagesgang des Luftverkehrs zeigt ein breites Maximum am Tage und ein Minimum um 01:30 Uhr Ortszeit; das Verhältnis von Maximum zu Minimum liegt bei 2,4. Innerhalb einer Woche wird der stärkste Verkehr donnerstags und freitags erreicht, gefolgt von einem Minimum am Sonntag; die Spannbreite liegt bei 18 %. Der Flugverkehr hat seine größte Dichte im Spätsommer und die geringste während des Winters. Im Kataster ist die Unterbrechung des Luftverkehrs nach dem Flugverbot vom September 2001 verzeichnet. Das vorliegende Kataster bildet eine wertvolle Datenbasis für realitätsnahe Simulationen der atmosphärischen Auswirkungen des Luftverkehrs in der oberen Troposphäre.

1 Introduction

Air traffic is expected to increase globally by a factor of 5 or 6 between 1990 and 2050 with a commensurate rise in emissions and contrails that may significantly affect air quality and climate (IPCC, 1996). Some of the aircraft exhaust effects, especially those impacting contrail and cirrus clouds, are still highly uncertain requiring more detailed research to more accurately prognosticate the climatic impact of enhanced commercial fleets. Contrail formation, growth, and dissipation and their optical properties are highly dependent on aircraft engine type, and the temperature, humidity, and wind speed and direction at flight altitude. The contrail-cirrus radiative effects, which ultimately affect the average state of the atmosphere, depend on the underlying conditions (surface temperature and albedo), the contrail optical properties, air traffic density and altitude, and the time of day when the contrails are formed. The formation and dispersion of some chemical species formed from aircraft exhaust also depend on the time of day when they are formed. Thus, to accurately assess current air traffic effects and future flight scenarios, it is necessary to simultaneously know the meteorological state and the distribution of flights at a given location and time. This paper addresses the latter need for the contiguous United States and southern Canada with a focus on the upper tropospheric portions of commercial flights.

The release of United States of America (USA) near-real time air traffic control information to the commercial sector during the late 1990's made the collection of more refined flight path data much easier than before. This report documents the collection, reduction, analysis, and availability of commercial flight information taken above 25,000 ft over the continental USA, southern Canada, northern Mexico, and adjacent waters since late 2000. The result of the analysis is a flight track database that can be easily accessed and used by re-

*Corresponding author: Patrick Minnis, MS 420, NASA Langley Research Center, Hampton, VA, 23681, USA, e-mail: p.minnis@nasa.gov

searchers focusing on the atmospheric effects of high-altitude aircraft.

2 Data and analysis

Commercial flight information over the USA from FlyteComm, Inc. has been purchased, downloaded in real time over the internet, and archived at NASA Langley Research Center (LaRC) since September 2000. The raw data comprise reports of flight number, aircraft type, download time, latitude, longitude, altitude, heading, destination and origination locations, speed, and departure and arrival times. These reports are updated every minute for all air traffic within range of the land-based air traffic radars and every 10° of longitude (∼30 minutes) for transoceanic flights that are out of the radar range. FlyteComm ingests the real-time feed from the USA Federal Aviation Administration (FAA) database and reformats it for commercial use. The FAA database includes, at a minimum, all USA Instrument Flight Rules (IFR) flights for which it is responsible (i.e., non-military), all IFR flights monitored by the Transport Canada Aviation Group, all IFR air traffic within several hundred miles of the USA borders, and all transoceanic IFR air traffic that originates or terminates in Canada or the USA. The FlyteComm data are then downloaded every 5 minutes by LaRC to a local computer hard drive. Because the data are taken in real time, interruption of any part of the process causes gaps. Sources of interruption include computer or power failures, full disks, or missing operators. The archived database, therefore, should be viewed as a sample rather than a complete representation of non-military flight traffic.

Daily data files, constructed of all flight reports within a 24-hour period, were sorted by flight ID and sub-sorted in turn by departure airport, arrival airport, and time. Data were qualified by eliminating reports that represented pending flights, had an altitude below 25 kft (7.6 km) or above 49.2 kft (15 km), had a location outside the analysis domain (20°N–50°N and 60°W–135°W), or exactly duplicated another line. Single flights were identified as unique combinations of flight ID, departure airport, and arrival airport. When a gap of more than 5 hours was found for one combination of flight ID, departure airport, and arrival airport, the data following the gap were identified as a different flight. This type of gap was usually cause by flights that traverse 0000 UTC. Flights having the airplane in two places at the same time were eliminated. The analysis grid is defined by longitude steps of 1°, latitude steps of 1°, and altitude steps of 1 km from 7 to 15 km, and was filled in time steps of 1 hour. The flight segments within each flight were sub-segmented by inserting additional interpolated points wherever a flight segment intersected the boundary between adjacent hypercubes of the analysis region. Simple summary statistics were then calculated for each hypercube.

Data covering an analysis month and the preceding year were collected in ensembles of 168-hour weeks with 0000–0100 UTC Sunday as the first hour. The mean and standard deviation were calculated for each track length of each hour of the ensemble (using only non-zero values), and any particular total track length value exceeding three standard deviations from the mean of the ensemble was considered to indicate an hour of invalid data. The process was repeated twice more, each time neglecting zero values or any hours identified as invalid, with remaining hourly data in the analysis month identified as valid. No data were removed or adjusted by this procedure. Each data file contains all of the data available for analysis. Estimated data validity is only provided as an aid to identify data useful for further analysis. More details of the qualification procedures are provided at the web site (see section 5).

The summary statistics for each month in Table 1, which shows the number of days in each month that have complete (full), partial, or no (empty) sampling, indicate that the best sampling occurred during late 2001 and 2002. No military or Visual Flight Rules (VFR) flights are available in this database. Not all IFR flights over Mexico are represented either. All commercial and private IFR flights over the USA and southern Canada should be included for those days noted as complete.

Flights remaining after qualification were used to develop the database, which is divided into two parts: linear and gridded. The former computes the node points for each flight track on a 1° latitude-longitude grid using interpolation along great circle arcs between each report. These standardized flight track positions constitute the linear database, which is in the form of one file for each day consisting of a series of flights, each with its own header describing the general flight characteristics and followed by a series of flight segments. The gridded database, provided in cell files, uses the segmented flight tracks to determine for each hour the number and total length of flights within a 1-km vertical range in a given 1° grid box. The linear dataset should be useful for detailed contrail simulation studies, while the gridded data should be more valuable for use in climate simulations.

Flights remaining after validation were used to produce the summaries given in the following sections. Hourly statistics were computed using only those hours with complete coverage, that is, no gaps were found in the record for a given hour. Daily means used only days with complete 24-hour coverage. Weekly statistics were computed by averaging the daily means for each day of the week using only those days with 24-hour coverage. Monthly statistics were computed using the weekly data weighted by the number of times a particular day of the week occurred during the month.

Table 1: Monthly sampling statistics for flight track database.

Month	2001 (number of days)				2002 (number of days)			
	Full	Partial	Empty	% Full	Full	Partial	Empty	% Full
Jan	8	7	16	26	16	5	10	52
Feb	0	0	28	0	16	7	5	57
Mar	16	4	11	52	23	7	1	74
Apr	15	14	1	50	28	2	0	93
May	12	16	3	39	22	7	2	71
Jun	14	16	0	47	29	1	0	97
Jul	17	13	1	55	27	4	0	87
Aug	15	14	2	48	28	3	0	90
Sep	22	6	2	73	25	4	1	83
Oct	22	5	4	71	23	8	0	74
Nov	16	6	8	53	17	11	2	57
Dec	25	5	1	81	11	15	5	35
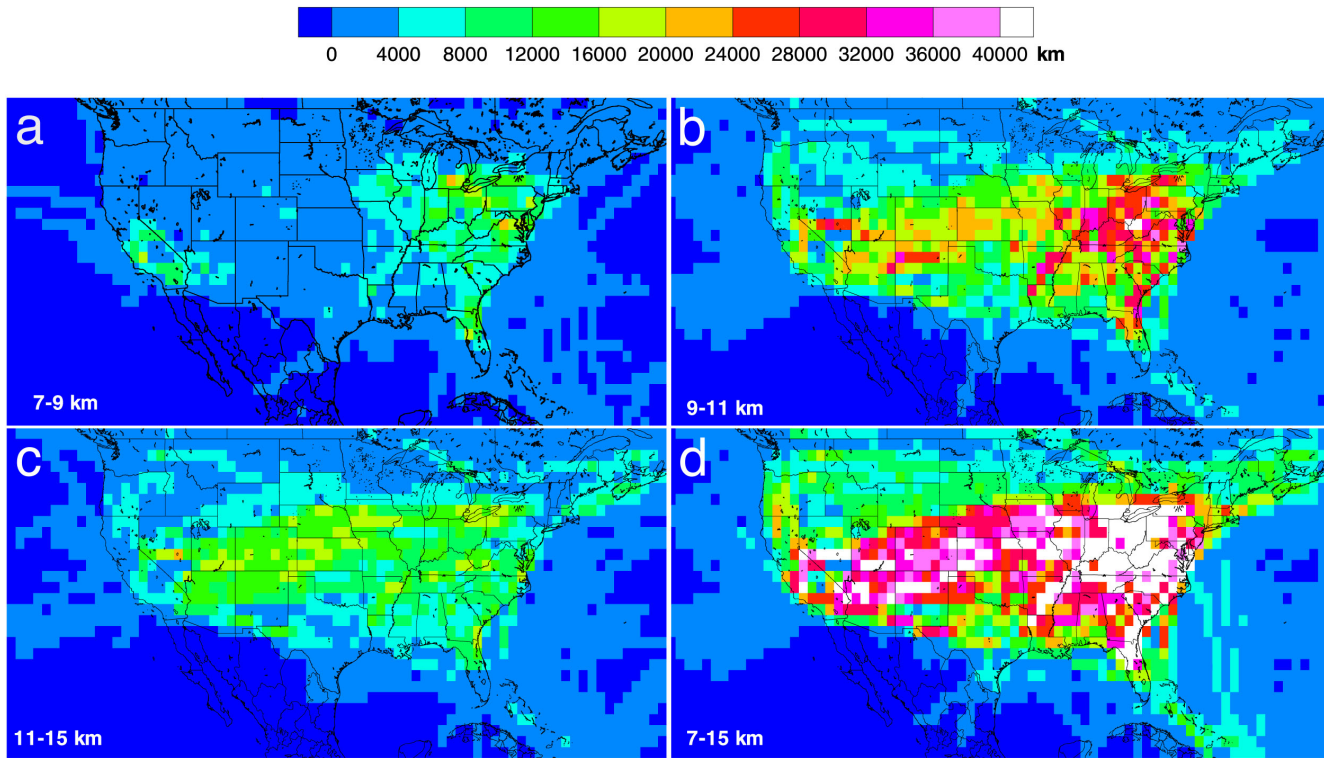
Figure 1: Cumulative flight lengths for 1° regions, 10 September 2001.

3 Results

The main parameters of interest for this database are the number of flights and the cumulative flight length (CFL). CFL can be computed as the sum of all flight segment lengths within a given box or domain for a selected time period, usually 1 hour.

3.1 Spatial variability

The 1° CFL distribution for 10 September 2001 is sliced by altitude ranges in Fig. 1. The maximum CFL in a single 1° box in the lowest altitude range is only ∼22,000 km over southern Michigan and northern Virginia. Most CFLs within that layer are less than 4000 km with many flights occurring in southern California,

northern Florida, and the Midwestern USA. Flights at these lower levels consist of portions of longer legs near the terminals or of the maximum altitude segments of short-distance commuter flights. The maximum CFL between 9 and 11 km exceeds 40,000 km over parts of eastern Ohio and western West Virginia. In addition to the large area of heavy traffic over the northeastern USA, a relatively dense traffic lane is evident over the Atlantic coast south of New York City. These flights are generally mid-range or longer flights. The maxima over the western USA are found over Nevada and the border between New Mexico and Arizona away from the large hubs at San Francisco and Los Angeles. The CFLs at the highest altitudes (Fig. 1c), generally consisting of long distance flights, are greatest over Lake Erie, the central
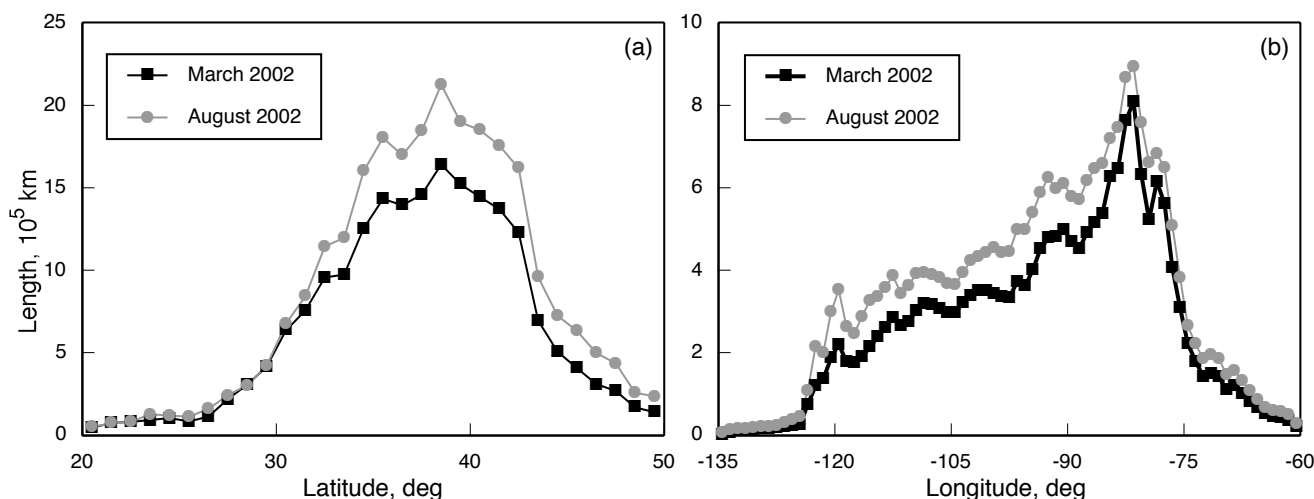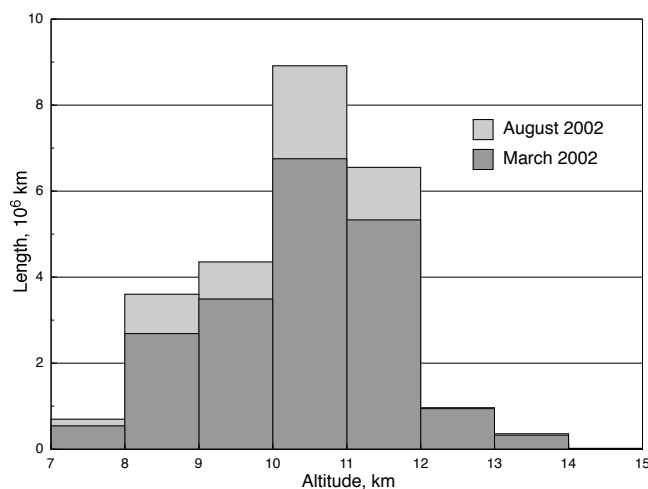
Figure 2: Mean daily cumulative flights lengths as a function of (a) latitude and (b) longitude over USA.

Great Plains, southern Utah, and coastal north Florida. The combined levels in Fig. 1d yield a large number of regions with total CFL > 40,000 km. These include much of the northeastern USA exclusive of New England, the Atlantic coast, central Great Plains, the Southwest, and lower Mississippi Valley. The few USA areas with total CFL < 4,000 km are found along the western Canadian and central Mexican borders. Data over interior Mexico is unreliable because there was no direct source of Mexican flight traffic information in the FlyteComm database.

The latitudinal and longitudinal variations of the mean daily CFLs are summarized in Fig. 2. Peak traffic occurs between 38°N and 39°N with a secondary maximum around 35.5°N (Fig. 2a). A relative maximum (Fig. 2b) at 120°W corresponds to the southern California traffic followed in the eastward direction by a dip and a relatively steady increase to the overall maximum at 81.5°W. The traffic then tails off to 30,000 km at 60°W. More flights are evident over the Atlantic than over the Pacific. The shapes of the longitudinal and zonal mean CFL curves are very similar for both March and August suggesting more of a general increase in traffic during summer than a changing of the air traffic patterns. The maximum CFL for the entire domain occurs between 10 and 11 km followed by the layer between 11 and 12 km (Fig. 3). The March-August increase in air traffic is greatest between 10 and 11 km. The total daily CFLs for the domain are 19.6 and 25.1 million km during March and August, respectively.

3.2 Temporal variability

Figure 4 shows the distribution of hourly cumulative flight segment lengths over the analysis domain for 10 September 2001. In this figure, the lengths of all flight segments within a given 1° box for all altitudes above 7.6 km were summed over a given UTC hour interval to


Figure 3: Mean daily layer cumulative flights lengths over USA domain.

yield the total length of flight tracks through each box. The sequence of plots begins with 0–1 UTC. The flight density is greatest over the eastern USA at that time but eastern traffic quickly drops off after local midnight, leaving mostly western traffic. Total traffic reaches a minimum during 9–10 UTC, and then rapidly builds over the east before spreading west with the sun. The hourly cumulative flight lengths rise to over 5000 km in a few grid boxes from 13 UTC and traffic remains robust until the end of the day.

The diurnal cycle over the entire domain is summarized in Fig. 5 using monthly means from March and August 2002. Fig. 5a shows a rapid increase in flights after 1100 UTC to a peak around 1830 UTC during March with a secondary maximum around 2330 UTC. The minimum occurs near 0730 UTC. During August, the air traffic begins in earnest an hour earlier than during March because of a shift to daylight savings time. The maximum during August occurs at 2230 UTC. Overall,
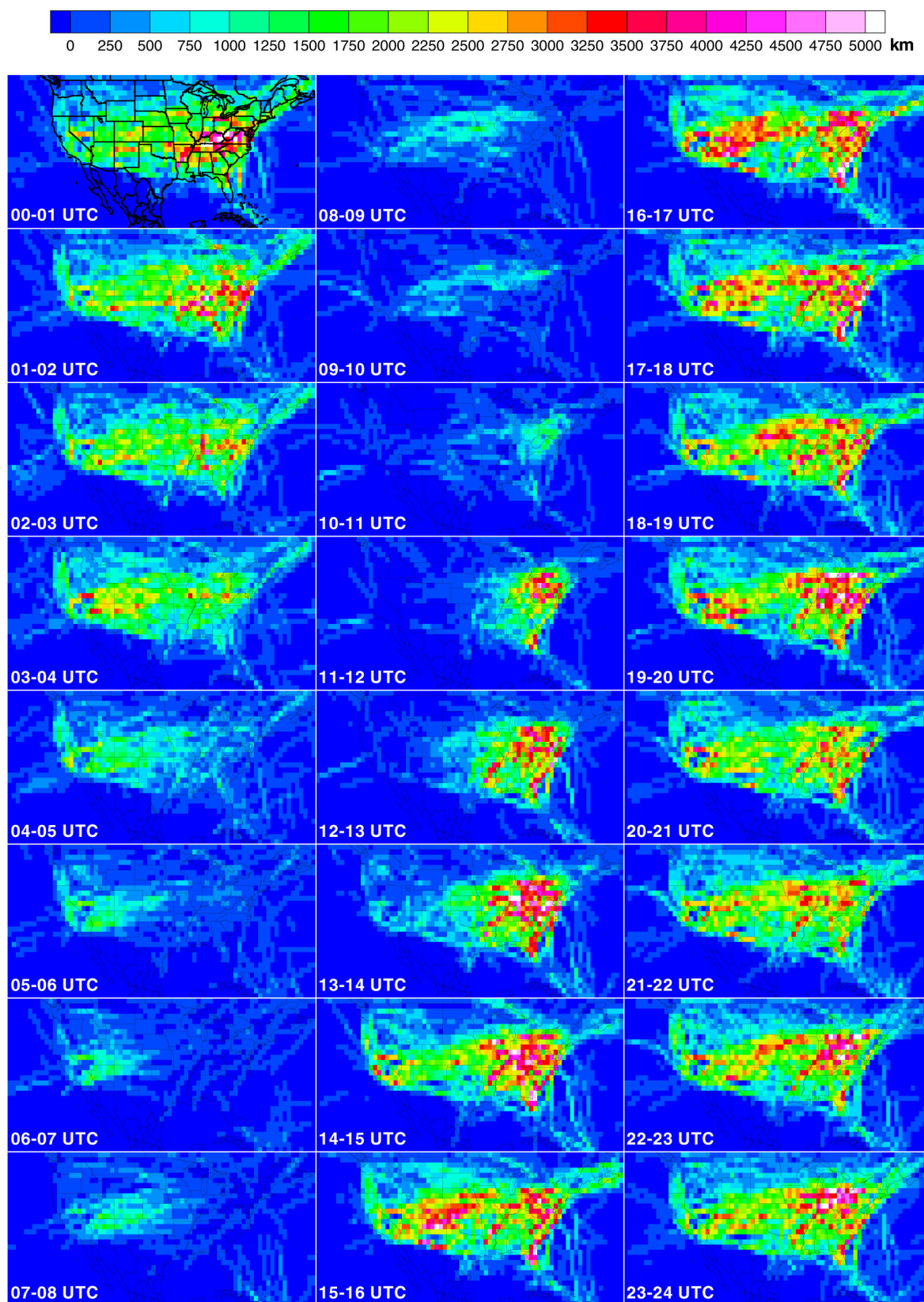
Figure 4: Hourly cumulative flight path length per 1° regional box, 10 September 2001.

the air traffic during August is ∼28 % greater than during March. The diurnal variation of air traffic shows no seasonal shift in local time (LT) with the daily cycle beginning in earnest after 0500 LT, reaching two maxima at 0930 and 1230 LT with a minimum at 0130 LT (Fig. 5b). Considering the hours 0600–1800 LT as day-
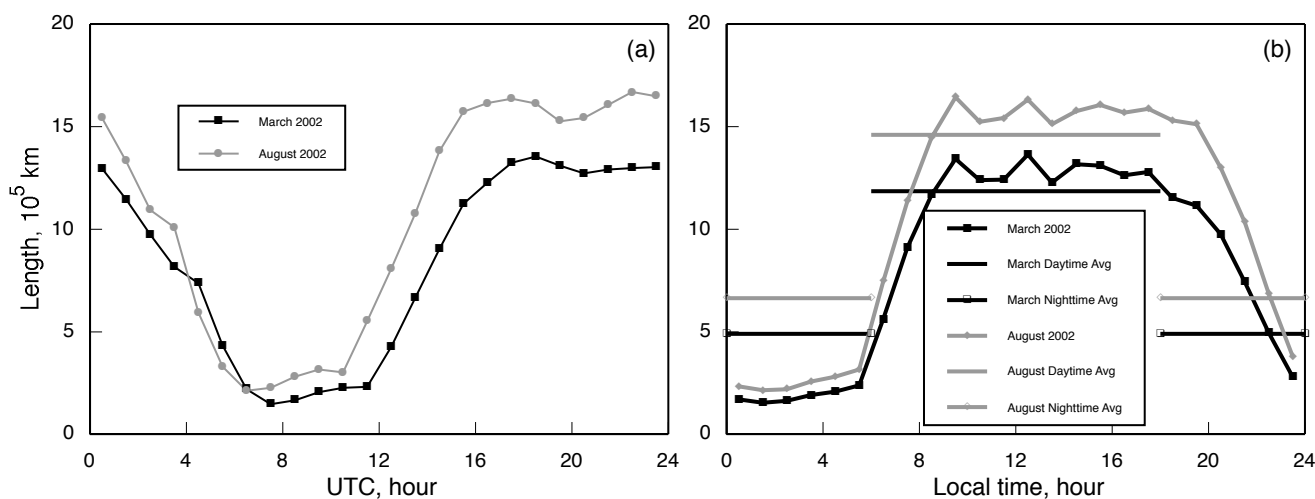
Figure 5: Mean hourly and day-night average cumulative flights lengths, (a) local time and (b) UTC.

time and the remaining hours as night, a rough estimate of the day-night ratio in cumulative flight length is 2.5 and 2.3, respectively, during March and August 2002. The true day-night ratio, valuable for calculating the relative shortwave and longwave radiative forcing by contrails, is more accurately determined for each month by considering the actual hours of daylight for each location and month. Such information can be easily computed from the gridded database.

The number of flights in the analysis domain has a distinct weekly cycle. The mean air traffic minimum occurs on Sunday. The mean number of takeoffs and landings rapidly rises to a plateau of 29,750 on Thursday and Friday and drops precipitously by 3,500 on Saturday. The range in the number of flights per day of week is typically around 5,000 except for the extreme minima occurring during certain holidays or as a result of the air traffic shutdown period after 11 September 2001. Relative to the weekly mean of 27,350 flights, the weekly range in the average daily number of flights is 18 %.

4   Discussion

While providing unprecedented air traffic detail, it should be noted that this dataset does not include flights below 25,000 ft, military air traffic, VFR flights, some air traffic over interior Mexico and Cuba, and some flights over open ocean (> 300 km from nearest USA or Canadian coast) that have no terminus in Canada or the USA. It is not known how many of the open ocean, Mexican, or Cuban flights have not been included because fo the limited nature of the original datasets. Development of a total inventory for this domain would require additional input. Data acquisition, while nominally continuous, was not complete except for a few months. The actions taken to account for the missing data when computing monthly or weekly means should have minimized the impact of the poor data. Data were initially

acquired from FlyteComm, Inc. every 5 minutes and the only time associated with each flight number was the local time on the acquisition computer. Thus, the actual flight time could be overestimated by up to 5 (greater over oceans) minutes depending on the time interval of reporting for a given flight. Despite these uncertainties in actual time, the data should provide a realistic representation of air traffic over the analysis domain.

Previous air traffic fuel usage data from USA sources were developed for 1992 (BAUGHCUM, 1996; GARDNER, 1998) and include military sources (METWALLY, 1995). A comparison of the fuel usage above 7 km from that earlier dataset (Fig. 1 in MINNIS et al., 2003) and the CFL distribution in Fig. 1d reveal some differences that may be due to the lack of military and foreign flights, changes in air traffic patterns since 1992, and the use of simple great circle routes between terminals in the 1992 dataset instead of actual flight tracks. For example, flights between California and Mexico, evident in the 1992 dataset, are absent in Figure 1d, while flights from Texas and the eastern USA to Acapulco and Mexico City (Fig. 1d) are not so well defined in the 1992 dataset. Air traffic over the Pacific is confined to a few narrow air lanes in the earlier dataset compared to the more diffuse air lanes in Fig. 1d. Despite many similarities between the two datasets, this air-lane versus diffuse distribution difference is apparent over much of the domain. For example, the maximum fuel expenditure in the 1992 dataset is confined to a latitude strip between 40°N and 42°N between Philadelphia, Pennsylvania and central Iowa with a few secondary maxima over Las Vegas, Nevada and southern California. In Fig. 1d, the maximum CFL covers a much larger area including the relatively narrow strip in the 1992 data. New secondary maxima are evident over northeastern Florida, eastern Arkansas, and eastern Kansas in Fig. 1d. Perhaps, additional tourist traffic and the expansion of overnight delivery services with major hubs at Memphis, Tennessee,
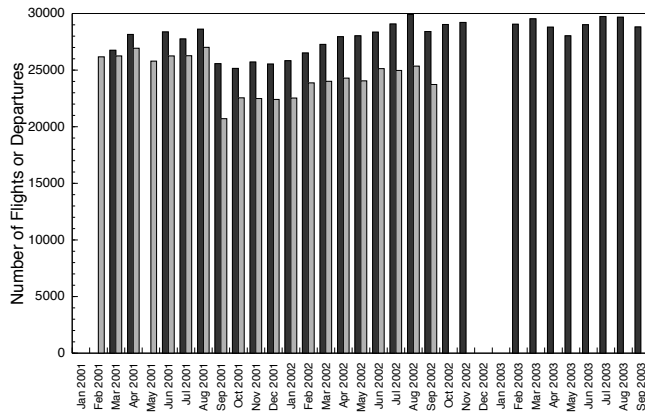
Figure 6: Mean number of daily flights above 7.6 km from FlyteComm data (black) and departures of commercial carriers from OAI database (shaded).

Wilmington, Ohio, and Louisville, Kentucky could have increased the traffic over the Midwest and Florida.

The seasonal variation in air traffic in the current dataset is somewhat similar to the annual cycle that can be inferred from the seasonal variation in North American high-altitude aircraft emissions from the 1992 inventory (Figs. 2–7, FRIEDL, 1997). The earlier data have a peak in NOx emissions during August followed by a 4 % drop into September, nearly constant values through December, a 9 % decrease to the January minimum, a gradual increase to May, and a rapid rise to the summer maximum. Figure 6 shows an initial August maximum during 2001 with flight frequency decreasing into October. The number of flights remains depressed until February 2002 when it begins increasing until August 2002. A 6 % decrease in flights during September 2002 is slightly offset by a 2 % rise into October. The flight frequency varied from 28,000 to 29,600 through September 2003. This irregular variability is influenced by a number of factors, including the 11 September 2001 terrorist attacks, which can account for part of the October 2001 – February 2002 lull. The anniversary of the attacks in September 2002 likely caused the dip during the same month.

The USA freight traffic, as expressed in overall ton-miles for scheduled and non-scheduled aircraft, increased almost monotonically by a factor of 2.7 between 1981 and 2000 with minor lulls during 1991 and 1999 (OAI, 2003). Similarly, the number of passenger originations increased by a factor of 2.4 during the same period. Both parameters decreased during 2001, especially the air passenger originations. As seen in Fig. 6, the overall number of scheduled and non-scheduled airport departures for large certified air carriers, which account for much of the high-altitude air traffic is 5–15 % less than the total number of flights in the current database (Fig. 6). The OAI (2003) dataset consisted of both small and large air certificated air carriers during 2000, but only large air carriers were included during 2002 (D. BRIGHT, personal communication, 2003). The transition in reporting practices occurred sometime between 2000 and 2002 accounting for the shift in the absolute number of reported departures during late 2001 compared to those determined from the FlyteComm data. The relative monthly variation in the OAI (2003) data is quite similar to the data compiled here, however, indicating that both datasets are tracking the seasonal variations. The FlyteComm data used here yield a larger number of flights because they are more inclusive than the OAI (2003) statistics having input from the Transport Canada Aviation Group.

The diurnal variation of flights is generally consistent with the available information on contrails over the continental USA. The unnormalized frequency of persistent contrails more than doubles between 0600 LT and 0800 LT when it reaches a maximum around 0900 LT (MINNIS et al., 1997, 2003). The average CFL also increases by more than a factor of 2 between 0600 and 0800 LT and reaches a peak at 0930 LT (Fig. 5). Mean CFL decreases very slowly during the daylight hours without a significant drop until after 1800 LT, while the contrail frequency also gradually decreases after 0800 LT before dropping at 1800 LT. The lack of contrail frequency data during the night precludes further comparison. However, the correspondence between contrails and flight lengths during the daytime suggests that the CFL data could be used to estimate the relative hourly frequency of contrail occurrence at night if it is assumed that the upper tropospheric temperature and humidity conditions are, on average, the same as those during the daytime.

5 Concluding remarks

This new database constitutes a different characterization of air traffic than previously available. It provides explicit flight paths and flight density in terms of cumulative flight length rather than fuel usage. Actual flight paths were used instead of great circle routings between terminals. Because the flight number and type of aircraft are recorded for each flight, it should be possible to estimate fuel usage and efficiency for each plane to provide additional input for atmospheric effects computations. Because of data dropouts, caution must be used when simulating the air traffic for a particular day. The statistics presented here summarize the dataset through September 2003. The effects of data dropouts were taken into account when producing the summary statistics. The database, which will be updated through August 2005, can be accessed via the World Wide Web at http://angler.larc.nasa.gov/flighttracks. Additional details concerning the data sampling and quality control are also available at the web site.

Air traffic over the USA is marked by distinct daily, weekly, and annual cycles that can affect evaluations of

aircraft effects on the atmosphere. When properly used, the dataset described herein should prove valuable for realistically simulating air traffic to address atmospheric effects at a variety of scales. In particular, the diurnal cycle of air traffic, which can influence many radiative and chemical processes dependent on solar energy, can be simulated with very high accuracy using this dataset. Other aspects of the database, such as the weekly cycle, may provide distinct signatures of air traffic in the climate record if the 2 years of data used here are representative of long-term averages.

Acknowledgements

References

BAUGHCUM, S.L., 1996: Subsonic aircraft emission inventories, In: Atmospheric Effects of Aviation: First Report of the Subsonic Assessment Project. – NASA RP-1385, 15–29.

FRIEDL, R..R. (Ed.), 1997: Atmospheric Effects of Subsonic Aircraft: Interim Assessment Report of the Advanced Subsonic Technology Program, NASA RP 1400, 168 pp.

GARDNER, R.M. (ED.), 1998: ANCAT/EC Aircraft Emissions Inventory for 1991/92 and 2015. Final Report EUR-18179, ANCAT/EC Working Group. – Defence Evaluation and Research Agency, Farnborough, United Kingdom, 108 pp.

IPCC, 1996: Climate Change 1995: The Science of Climate Change. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change. [HOUGHTON, J.T., L.G. MEIRA FILHO, B.A. CALLANDER, N. HARRIS, A. KATTENBERG, K. MASKELL (Eds.)] and WMO/UNEP. – Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 572 pp.

METWALLY, M. 1995: Jet Aircraft Engine Emissions Database Development-1992 Military, Charter, and Non-scheduled Traffic. – NASA CR-4684.

MINNIS, P., J.K. AYERS, S.P. WEAVER, 1997: Surface-Based Observations of Contrail Occurrence Frequency Over the U.S., April 1993 - April 1994, NASA RP 1404.

MINNIS, P., J.K. AYERS, M.L. NORDEEN, S.P. WEAVER, 2003: Contrail frequency over the USA from surface observations. – J. Climate **16**, 3447–3462.

OAI, 2003: Air Traffic Statistics and Airline Financial Statistics, Historical Air Traffic Data for 1981–2001. – U.S. Department of Transportation, Bureau of Transportation Statistics, http://www.bts.gov/oai/indicators/top.html.